

WHAT IS CLAIMED IS:

1. A translation apparatus for translating article information in a first language including an article body and a related headline into a second language, comprising:

a decision unit configured to discriminately identify the article body and the headline; and

a translation unit configured to translate the article body and the headline into the second language based on the identification result of said decision unit.

2. The translation apparatus according to claim 1,

wherein said decision unit determines whether the article information is registered as a news site, and discriminates the article body and the headline based on a decision algorithm corresponding to the news site if the article information is registered as a news site.

3. The translation apparatus according to claim 2,

wherein said decision unit discriminates the article body and the headline based on a decision

algorithm corresponding to a general news page if the article information is not registered as the news site.

4. The translation apparatus according to claim 1,

further comprising a phrase alignment processing unit configured to respectively extract a noun phrase from the headline and a noun phrase candidate from a head sentence of the article body, to count coincident words between the noun phrase and the noun phrase candidate, and to output the noun phrase and the noun phrase candidate as a phrase alignment result if a coincidence ratio based on the counted value is above a threshold.

5. The translation apparatus according to claim 4,

wherein said phrase alignment processing unit detects continuous capital letters as an abbreviation from the headline, detects continuous words each including an initial capital letter as a formal phrase from the head sentence of the article body, compares the continuous capital letters with a set of combined capital letters extracted from the continuous words, and additionally counts as one

coincident word if the continuous capital letters coincide with the combined capital letters.

6. The translation apparatus according to claim 5,

wherein said phrase alignment processing unit converts the abbreviation in the headline to the formal phrase if the coincidence ratio based on the counted value is above the threshold.

7. The translation apparatus according to claim 4,

wherein said phrase alignment processing unit extracts a noun phrase at the end of the headline, extracts a predetermined pattern including a noun phrase from the head sentence of the article body, and identifies the noun phrase of the headline as an information source part if the noun phrase of the headline and the noun phrase of the predetermined pattern are included in the phrase alignment result.

8. The translation apparatus according to claim 1,

further comprising:

a first language-second language parallel corpus configured to store a plurality of articles

in the first language and corresponding translated articles in the second language;

a similar article retrieval unit configured to retrieve an article in the first language and a corresponding translated article in the second language from said first language-second language parallel corpus, the article in the first language describing subject matter similar to the article information for translation, and

a target word information extraction unit configured to extract target word information from the translated article in the second language, and to supply the target word information to said translation unit.

9. The translation apparatus according to claim 8,

wherein said similar article retrieval unit calculates a similarity degree between frequency of each word in the article information for translation and frequency of each word in each article of the first language stored in said first language-second language parallel corpus, and selects one article of the first language of which the similarity degree is maximum from said first language-second language parallel corpus.

10. The translation apparatus according to claim 9,

wherein said similar article retrieval unit lowers a weight of decision of similarity degree for at least one word of a proper noun, a date and a quantity.

11. The translation apparatus according to claim 8,

wherein said first language-second language parallel corpus correspondingly stores each word in each article of the first language and each target word in the translated article of the second language.

12. The translation apparatus according to claim 1,

wherein said translation unit translates the headline of the first language into the second language by referring to a special rule for translating the headline.

13. A translation method for translating article information in a first language including an article body and a related headline into a second

language, comprising:

discriminately identifying the article body  
and the headline; and

translating the article body and the headline  
into the second language based on the identification  
result.

14. The translation method according to claim  
13,

further comprising:

determining whether the article information is  
registered as a news site; and

discriminating the article body and the  
headline based on a decision algorithm corresponding  
to the news site if the article information is  
registered as a news site.

15. The translation method according to claim  
14,

further comprising:

discriminating the article body and the  
headline based on a decision algorithm corresponding  
to a general news page if the article information is  
not registered as a news site.

16. The translation method according to claim

13,

further comprising:

respectively extracting a noun phrase from the headline and a noun phrase candidate from a head sentence of the article body;

counting coincident words between the noun phrase and the noun phrase candidate; and

outputting the noun phrase and the noun phrase candidate as a phase alignment result if a coincidence ratio based on the counted value is above a threshold.

16, 17. The translation method according to claim

further comprising:

detecting continuous capital letters as an abbreviation from the headline;

detecting continuous words each including an initial capital letter as a formal phrase from the head sentence of the article body;

comparing the continuous capital letters with a set of combined capital letters extracted from the continuous words; and

additionally counting as one coincident word if the continuous capital letters coincide with the combined capital letters.

18. The translation method according to claim 17,

further comprising:

converting the abbreviation in the headline to the formal phrase if the coincidence ratio based on the counted value is above the threshold.

19. The translation method according to claim 16,

further comprising:

extracting a noun phrase from the end of the headline;

extracting a predetermined pattern including a noun phrase from the head sentence of the article body; and

identifying the noun phrase of the headline as an information source part if the noun phrase of the headline and the noun phrase of the predetermined pattern are included in the phrase alignment result.

20. The translation method according to claim 13,

further comprising:

storing a plurality of articles in the first language and corresponding translated articles in



the second language in a first language-second language parallel corpus;

retrieving an article in the first language and a corresponding translated article in the second language from the first language-second language parallel corpus, the article in the first language describing subject matter similar to the article information for translation; and

extracting target word information from the translated article in the second language.

21. The translation method according to claim 20,

further comprising:

calculating a similarity degree between frequency of each word in the article information for translation and frequency of each word in each article of the first language stored in the first language-second language parallel corpus; and

selecting one article of the first language of which the similarity degree is maximum from the first language-second language parallel corpus.

22. The translation method according to claim 21,

further comprising:

lowering a weight of decision of similarity degree for at least one word of a proper noun, a date, and a quantity.

23. The translation method according to claim 20,

further comprising:

correspondingly storing each word in each article of the first language and each target word in the translated article of the second language in the first language-second language parallel corpus.

24. The translation method according to claim 13,

further comprising:

translating the headline of the first language into the second language by referring to a special rule for translating the headline.

25. A computer program product, comprising:

a computer readable program code embodied in said product for causing a computer to translate article information in first language including an article body and a related headline into a second language, said computer readable program code having:

a first program code to discriminately identify the article body and the headline; and

a second program code to translate the article body and the headline into the second language based on the identification result.

205210 78245001